

## Social Collaboration Analytics for Enterprise Collaboration Systems: Providing Business Intelligence on Collaboration Activities

Florian Schwade  
University of Koblenz-Landau  
Institute for Information Systems Research  
Faculty of Computer Science  
Germany  
[fschwade@uni-koblenz.de](mailto:fschwade@uni-koblenz.de)

Petra Schubert  
University of Koblenz-Landau  
Institute for Information Systems Research  
Faculty of Computer Science  
Germany  
[petra.schubert@uni-koblenz.de](mailto:petra.schubert@uni-koblenz.de)

### Abstract

*The success of public Social Media has led to the emergence of Enterprise Social Software (ESS), a new type of collaboration software for organizations that incorporates “social features”. Surveys show that many companies are trying to implement ESS but that adoption is slower than expected. We believe that in order to understand the issues with its implementation we need to first examine and understand the “social” interactions that are taking place in this new kind of collaboration software. We propose Social Collaboration Analytics (SCA), a specialized form of examination of log files and content data, to gain a better understanding of the actual usage of ESS. Our research was guided by the CRISP-DM approach. We first analyzed the data available in a leading ESS. Together with leading user companies of this ESS, we then developed a framework for Social Collaboration Analysis, which we present in this paper.*

### 1. Introduction

The increasing use of Social Media and their “social software features” in private life has changed the way people communicate and exchange information and has stimulated expectations on the side of employees regarding the use of similar software features in their workplace [1]. The increased demand for *socially-enabled collaboration software* has given rise to a new type of business software, which is called Enterprise Social Software (ESS) by researchers [2], and which has become an integral part of companies’ Enterprise Collaboration Systems (ECS).

Large software vendors such as IBM, Microsoft and Atlassian have launched new software products (e.g. IBM Connections, Yammer or Atlassian Confluence), which provide social software features in an integrated platform. Such integrated collaboration systems combine collaboration and social features in one platform and establish uniform access for employees

(single-sign-on, uniform user interface). *Integrated collaboration systems* also provide activity logs that allow us to analyze multiple forms of collaborative work in the digital workplace in a new way, which presents a great opportunity for researchers to explore and better understand the collaboration activities that are going on in companies.

However, whilst these integrated ECS record all activity in log files, in their current form they provide only very rudimentary standard functionality for the *analysis of activity logs*. Examples of standard reports provided by these systems are the “number of active users” or the “number of workplaces” on the platform. This information is not sufficient to demonstrate the business value that an ECS provides to a company [3].

*Social Analytics* is an emerging research domain in the context of Social Media that addresses the user engagement and audience sentiment evolution in Social Networks [4]. First attempts have been made to transfer such approaches to ESS, e.g. in the form of *network analysis* (responding to questions such as “how are people related?”), *community analysis* (“what kind of communities exist or form?”) and simple statistics on the *intensity of use* (“what percentage of the user base is actually using the system?”). The content of integrated systems such as IBM Connections or Atlassian Confluence allows a more detailed and more sophisticated analysis but there is a lack of tools to systematically analyze log files and visualize the results.

From our literature analysis and a primary survey of companies that are using socially-enabled ECS [5] we identified three existing challenges for *Social Collaboration Analytics*:

1. Understanding the log files and tables in the database of the collaboration software
2. Knowing what can be analyzed and what information can be extracted

3. Programming the queries and finding ways of presenting the results (reports containing tables and graphics)

In our paper, we address these three questions. We applied a *multi-stage design science research (DSR) process* [6] to develop and test an approach to *Social Collaboration Analytics (SCA)*. The development of the approach was guided by the phases of the CRISP-DM cycle [7]. We first analyzed and deciphered the database structure of a leading *integrated ECS* to develop our *data understanding*. We then performed a needs analysis with end users (who provided the *business understanding*), in our case a selected number of companies using the same collaboration software. We then developed prototypical reports (*modeling*) and evaluated them with our user companies in a focus group meeting (*evaluation*). The results of our work are presented in this paper.

## 2. Literature Review and Definition of Terms

**Social Collaboration Analytics.** We use the term “Social Collaboration Analytics” to describe the approach for analyzing and displaying collaboration activity of users in socially-enabled collaboration systems. Social Collaboration Analytics contains important elements from Web Analytics. Web Analytics can be described as “the whole process from data gathering to recommendations for website redesign” [8, p. 4]. Web Analytics uses concepts from Web Content Mining, Web Structure Mining and Web Usage Mining [8]. Social Collaboration Analytics makes use of these concepts but has its focus on the analysis of collaboration activities described in the core of Williams’ *8C Model for Enterprise Information Management* (communication, cooperation, coordination and content combination) [9]. The main difference between Web Analytics and Social Collaboration Analytics is that WA examines user activity with the final goal of optimizing the Website for a *single user* (1:1) while SCA has the *interaction patterns of many users* (m:n) in its focus. We argue that the analysis of *collaborative activities* requires new data sources, methods and data dimensions. Therefore, Social Collaboration Analytics can be seen as an extension of traditional Web Analytics with a focus on the analysis of usage, communication and interactions that occur in an Enterprise Collaboration System. For a better understanding of the field of Social Collaboration Analytics, we will make our understanding of the terms Social Media, Enterprise Social Software (ESS), Enterprise Social Network (ESN) and Enterprise Collaboration Systems (ECS) explicit in the next

paragraphs of this paper (cf. Figure 1).

**Social Media** are (public) platforms for social interaction and information exchange. They are characterized by their openness (any interested person can register and use the platform) and by their ownership (they are usually provided by a company who owns the platform and in most cases, by means of terms and conditions, also the user generated content). In Social Media, people gather voluntarily in their free time to chat, exchange ideas, to play together and, most importantly, share information (photos, films, files).

**Social Software** is a software category that provides social features such as social profiles, follows, likes and different forms of content (e.g. chat, blog, wiki, bookmarks, files) [10]. Social software usually provides new capabilities of content aggregation such as activity streams. The social features that emerged in Social Media triggered the development of specialized social software for companies.

**Enterprise Social Software (ESS)** is the term used for collaboration software that incorporates social media functionality such as social profiles, microblogging [11], wikis, blogs and file sharing. [1], [2], [12]. A combination of ESS and classical groupware (e.g. E-Mail, group calendars) can be combined to build an Enterprise Collaboration System.

An **Enterprise Collaboration System (ECS)** is a means of electronically supporting collaboration in a company. ECS support all areas of collaboration such as information and content sharing, communication, cooperation and coordination as described in the 8C Model for Enterprise Information Management [9]. ECS come in two different configurations: as *integrated systems* which provide multiple applications (modules) under a uniform user interface or as a *portfolio*, which combines various different applications from, in some cases, different software manufacturers [1], [13].

**Enterprise Social Network (ESN)** is a term for the structures that form between users of an Enterprise Collaboration System. Social profiles, which contain all information about a person including role, expertise and contact information, form the basis of an ESN. The links between the users are established with the help of

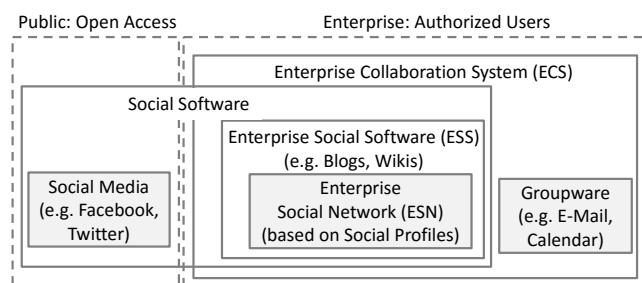


Figure 1: Terminology used in this paper

software features such as “invite to my network”, “follow” and “share content” [2], [14].

### 3. Research Design: Three Stage Approach

Our research for the development of an SCA framework was organized in three phases and guided by the CRISP-DM approach for data mining [7]. The phases contained the following activities:

1. Phase 1: Understanding the integrated ECS (system examination)
2. Phase 2: Business needs analysis (focus group)
3. Phase 3: Prototypical reports and evaluation (workshop)

For the development of our SCA framework we followed the steps of the CRISP-DM (CRoss-Industry Standard Process for Data Mining) cycle [7]. As shown in Figure 2, CRISP-DM suggests six steps for the pro-

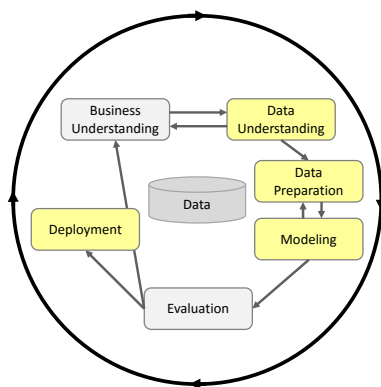


Figure 2: CRISP-DM Cycle

cess of data mining. In this paper, we describe the phases of *data understanding* and *data preparation*. We also show an example of a possible *deployment*. The actual framework contains the dimensions for the

development of queries (*modeling*). We illustrate the framework with an example of a possible *deployment*.

**Research phases.** Figure 3 shows the research phases and their activities. *In the first phase*, we defined our research objectives for the development of a framework for Social Collaboration Analytics (SCA). We performed a structured literature review searching for terms, definitions and current practice in the broader area of Web Analytics, Social Analytics and Data Mining. We selected a leading industry solution of an Enterprise Collaboration Systems (IBM Connections) for our practical examination. We analyzed the databases of the ECS and documented the data scheme.

*In the second phase*, we engaged in a multi-client University-Industry Collaboration [15] in order to make our research relevant for practice. We selected companies that are using the selected ECS and ran a focus group. In the focus group, we presented the data scheme and invited the practitioners to think about questions and desired reports that our framework should provide. In this phase, we also examined the academic literature for existing methods, techniques and dimensions that could help with our Social Collaboration Analytics. Some of the findings are presented in the section containing the literature review below. The findings from the focus group provided us with desired questions and outcomes for our SCA framework.

*In the third phase*, we developed the SCA framework and generated some exemplary reports. These were presented and discussed at a workshop with the user companies. The findings from the workshop provided us with the first evaluation of our approach. We are currently applying the framework creating further reports and searching for additional ways for displaying/presenting the SCA results. We will describe the work of the three phases in more detail in the next sections.

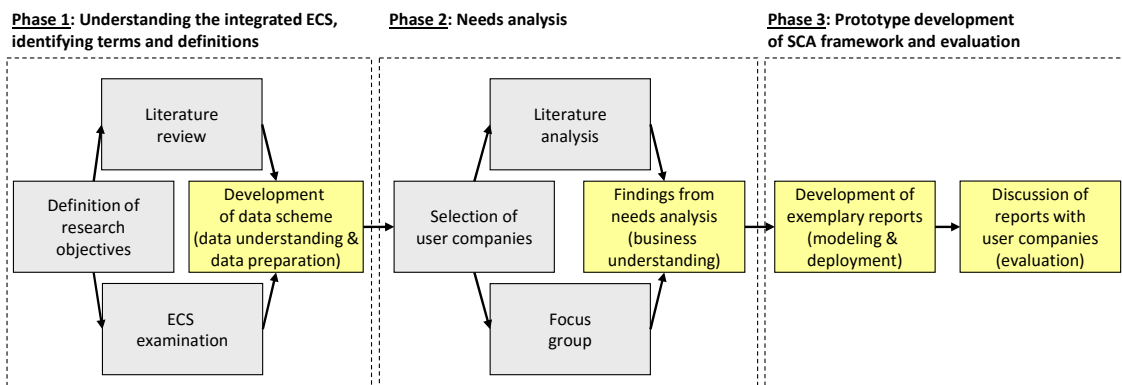


Figure 3: Research steps

### 3.1 Phase 1: Literature Review and Examination of the ECS

In this section, we report on the outcome of phase 1, the findings from our literature review and the data scheme.

#### 3.1.1 Literature Review

Behrendt et al. [14] identify four *data dimensions* for the analysis of ESN: (1) activities, (2) content, (3) relations and (4) experiences. The authors describe the first dimension, *activities*, as *usage* data. This type of data can be collected either by analyzing log files or by exporting relevant data directly from the systems database. Web Analytics tools can be used to analyze usage data. The *content* is referred to as *user-generated* data, such as status updates and blog posts, enable researchers to draw conclusions on *why* users have visited the platform. For the analysis of user-generated data, content analysis, sentiment analysis, text mining or genre analysis can be applied. The third dimension, *relations*, is described as *structural* data. In ESN structural data occurs when users create connections. Graphs can be used to analyze and visualize structural data. Behrendt et al. [14] postulate that structural data can be exported easily from the systems, however, they note that the analysis is challenging. *Experiences* are *reported data* by users. Reported data can be gathered by means of interviews or questionnaires. Behrendt et al. [14] observe that a huge effort and resources are required for the collection of such information. Content analysis can then be applied to analyze the reported data. Furthermore, the authors argue strongly that a mixed methods approach is required to perform a comprehensive analysis of user interactions in Enterprise Social Software.

In the context of *measuring the success of ESS*, Herzog et al. [3] introduce four methods with metrics for the usage analysis in ESS. *Content* and *usage analysis* can be conducted by measuring the number of adjusted ideas, the intensity of collaboration and the degree of cross-linking. The authors suggest *database queries/log file analysis* to measure ECS usage. For database queries and log file analysis a range of metrics that are mostly known from Web Analytics are suggested such as the number of posts, visits, edits, created pages per day or the average time per user per visit. However, the authors do not specify a process that describes how to develop queries or how to conduct a log file analysis. Similar

to Behrendt et al. [14] they mention that *user interviews* and *user surveys* can be conducted to gather additional information. Suggested metrics for user interviews are for example the usage behavior and review of the used tool. For the surveys, Herzog et al. suggest metrics such as user satisfaction and the frequency of use.

We argue that previous research is limited in several ways. The literature consistently points to the extensive work that would be necessary to develop further and more complex methods of analysis and researchers seem to shy away from the development because of this effort. Some authors even postulate that the current features based on traditional Web Analytics that are implemented in current ECS *are sufficient* for analysis. Herzog et al. [3] follow a similar line of argumentation. They state that “Nearly every Social Software platform provides analysis tools with which it is possible to track the activity of users in the form of traffic, page views and number of users, logins, or blog posts. Furthermore, this method meets the need for quantifiable success metrics of an IT manager” [3, p. 5]. We do not agree with this assessment. The results of our workshops with user companies clearly showed that current metrics tools in ESS are not satisfactory. Managers wish to see analyses that are more complex and meaningful in order to be able to assess the success and usage of their systems.

#### 3.1.2 Data Scheme: Data Understanding and Data Preparation

In order to develop the general data scheme for our Social Collaboration Analytics we first had to develop an understanding of the data available in the ECS. We examined the databases and tables contained in IBM Connections (which is deployed on a WebSphere ap-

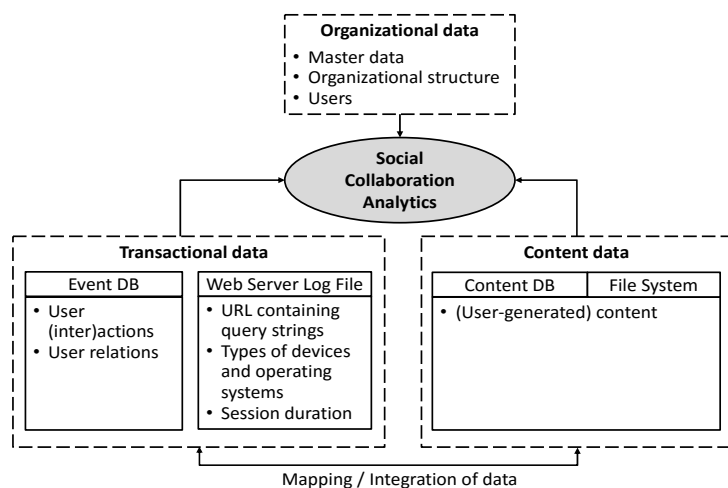


Figure 2: Data sources for Social Collaboration Analytics

plication server). This process took several months because the different applications (modules) of the software are based on multiple databases and contain hundreds of tables and the database structure is undocumented by the software vendor. Our findings showed that there are three important data sources that need to be considered for our SCA: (1) organizational data, (2) transactional data and (3) content data. Figure 2 shows the three data types and their characteristics.

**Organizational data** describes the organization with regard to its structure and its employees (users of the ECS). Technical implementations are the company-wide LDAP with its people, affiliations and their associated access rights.

**Transactional data** is the most important source for SCA and is stored in two different databases. The *event database* stores all actions that are initiated by the users. It contains records including the action type, the content type, the user and a time stamp. User (inter)actions and the relations between users can be derived from this source as secondary data. The event database is the log file of all user activity of an integrated ECS. It comprises multiple tables; the main event table and several lookup tables that provide further information on the keys used in the event log. In natural language, a record in the event table would read as follows:

*User with ID 26 created a content item with ID 106 (i.e. a blog entry) in the community with ID 1030 at 10:30 a.m. on the 7<sup>th</sup> of June 2016.*

ID	ITEM_UUID	USER_ID	EVENT_OP_ID	COMMUNITY_ID	EVENT_TS

ID	USER_ID	TITLE	TEXT	PUBTIME	WEBSITEID	HITCOUNT

**Figure 3: Excerpts from the event table and a content table**

Figure 3 depicts an excerpt with the most important information that is contained in such a record. With the help of the lookup tables, it is possible to derive further information from the event table such as the content type of the object or the name of the community in which the action occurred.

The second source for transactional data is the *log file of the webserver*. The records in the log file allow extraction of information for example from the URL that contains the query string, the device type and the operating system that are used to access the system and the session duration. This information is not stored in the event log of the ECS so the log file of the webserver (in our case an Apache http server) has to be analyzed to retrieve this information.

The **content data** is the third data source. Content data can be either stored in a database (e.g. forum posts, status updates etc.) or in the file system (e.g. uploaded files or attachments). Therefore, also the file system of an ECS needs to be included for the SCA.

As transactional data and content data are stored in two different data sources, the mapping between trans-

actional data and content data is an important issue for Social Collaboration Analytics. To achieve this, the Universally Unique Identifiers (UUID) that identify a content object need to be retrieved. Each content object is associated with a UUID that is valid across all data sources. Figure 3 shows excerpts from the event table and from a table of the content database. The figure also indicates how references between transactional data and content data can be established. To improve the comprehensibility of the figure, we limited the information to the most important fields that are required for Social Collaboration Analytics.

Each record in the event table has a unique ID. The ITEM\_UUID is the unique identifier of a content object in the system and provides a link to the respective record in the content table. This UUID is valid across all data sources. The USER\_ID stores the user who is responsible for the logged action. The EVENT\_OP\_ID specifies the type of action. There are multiple types of action such as create, delete, edit, follow and many more. In the case of IBM Connections, there are more than 50 types of action that can occur. The COMMUNITY\_ID specifies the community in which the traced event occurred. The column EVENT\_TS contains the timestamp of the record and thus documents the exact time and date at which the action occurred.

A record in a content table contains similar but slightly different information. The figure above depicts an excerpt of the information of the table for blog entries. As before, each content object has a unique ID. Further, like in the event log, the USER\_ID of the user who created the content is stored. Each blog post has a TITLE and a TEXT. The actual content of the blog post is stored in the field TEXT. The field PUBTIME documents the time at which the blog post was *(last) published*, which means that the content and the previous timestamp are continuously overwritten and it is not possible to determine the points in time when this content was changed. This information is only contained in the event table. The WEBSITEID specifies the blog with which this *blog post* is associated. Finally, the HITCOUNT provides information on how often a blog post has been visited to date. A record from the content table reads as:

*“User with ID 26 published the blog post ‘Hello World!’ in the blog with ID 1337 at 10:30 a.m. on the 7<sup>th</sup> of June 2016”.*

The example shows that it is not possible to construct (inter)action logs from the content table as the original content item is overwritten when it is updated – except for modules that provide version control. In order to retrieve the blog to which this specific content belongs, a join with a lookup table has to be performed. We found that even in this case not all required information about the action can be retrieved from the content database. Therefore, it is necessary to join the data from the *event table* and the *content table* for meaningful Social Collaboration Analytics. In some cases, this proves to be difficult but in most cases the joins can be established with the help of matching USER\_ID and/or the ITEM\_UUID of the content object as these are the two variables that are unique across all data sources.

### 3.2 Phase 2: Creating the Necessary Business Understanding (Needs Analysis)

As mentioned before, the second phase of our research was aimed at ensuring the practical relevance of our research [16]. We performed a focus group with 11 practitioners from companies that are participating in a University-Industry Collaboration initiative named IndustryConnect [5]. All members of this initiative are user companies and early adopters of ECS. First, we presented the results of our data examination (the preliminary data scheme) in order to enable the participants to assess the available information that could potentially be used for analysis. We then invited them to develop and present questions that they would like to pose regarding the use of their collaboration system and the interaction of their employees. The findings from the focus group confirmed the importance of our SCA research and demonstrated that managers of collaboration platforms need sophisticated means to analyze collaboration activities. It was repeatedly pointed out by the participants that the analysis features that are currently available do not provide sufficient information for the assessment of user activity.

Furthermore, possible dimensions, which can be used for SCA were discussed. These findings provided us with the necessary input for the next phase in which we modeled the data and developed the SCA framework.

### 3.3 Phase 3: Modeling, Deployment and Evaluation

Figure 4 depicts the architecture of the integrated ECS. Although the architecture was derived from a specific ECS, we believe that it can serve as a representation of the data sources of integrated ECS in general because usually ECS consist of multiple modules and therefore follow a similar architecture. The names in the boxes indicate the names of the modules. The descriptions in brackets indicate the name of the database of each application. Each application (module) of the examined ECS stores its data in an independent database.

The following section provides an overview of the different modules that exist in the examined ECS. The *homepage* contains the individual news stream for each user. In the homepage all news that the user is subscribed to (e.g. changes to content, status updates etc.) are gathered and displayed [17], [18]. *Communities* are shared workspaces for teams [19]. They act as “containers” and provide the members of this community with selected applications (modules) of an integrated ECS (e.g. microblog, blog, Wiki). A *social profile* represents a user in an ECS. The profile contains information about the person (e.g. a photo, name, location) and can be used to identify employees with expertise in a specific domain [20], [21].

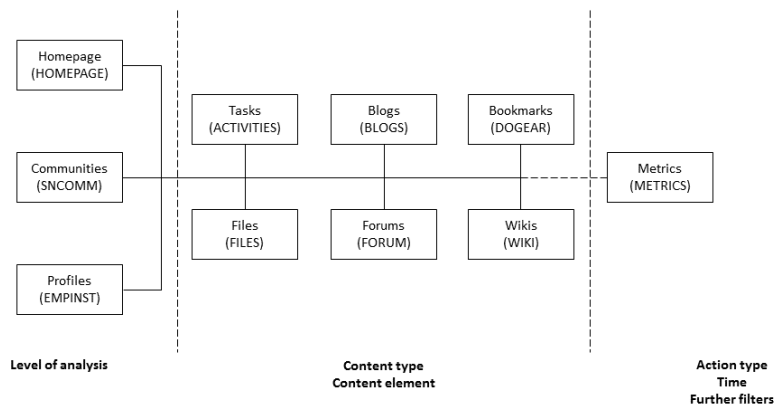


Figure 4: Architecture of an integrated ECS based on [43]

Integrated ECS are composed of different applications (modules), which we describe in this section. *Activities* (or shared tasks) provide users with an opportunity for shared task management. Task items can be created and users and deadlines can be assigned. *Blogs* are characterized by a chronological order of posts [22], [23]. Blogs entries usually contain text on a particular topic including embedded pictures, links and sometimes even file attachments. *Bookmark* applica-

tions enable users to create collections of bookmarks (URLs) which can be shared with other users [24]. *File* sharing applications enable users to share different types of files with other users [25]. In a *forum* discussion entries or questions can be posted which are usually dealing with a specific issue or problem. A sequence of such posts is called *thread* [26]. *Wikis* are collections of web pages, which are characterized by cross-links. The key feature of a Wiki is that each page can be edited while a complete history of the changes is kept. Therefore, Wikis are often used to build collaborative knowledge repositories with multiple authors [27]. Finally, the *metrics* application stores all actions that are initiated by the users and therefore collects the aforementioned event logs.

The content that is created in socially-enabled ECS is a special kind of content referred to as social business documents [22]. Social business documents are special for several reasons. Firstly, they are created and edited *collaboratively*. Secondly, social business documents are *compound documents* because they consist of multiple components such as the main text, *embedded elements* such as pictures and *attached elements* such as tags or comments. Thirdly, social business documents are inherently social because they show up in the activity stream of followers or subscribers and are linked to social profiles, which contain further information about their authors and thus provide rich background information (e.g. for an assessment of the quality of the content). This understanding of social business documents is an essential foundation for our SCA framework, which is proposed in the following.

As mentioned before, we developed the SCA framework based on the findings from the literature analysis and the input from the focus group. As shown previously, each record in an event log can be read as a sentence in natural language. The event log is the most important data source for Social Collaboration Analytics. The SCA framework is meant to assist the development of questions that can be answered with the help of the available data and is thus aligned with the architecture of the integrated ECS. The proposed SCA framework is depicted in Figure 5.

The items in the *six columns* guide the forming of questions. The selected content is used to form the predefined SQL statements. First, *the level of analysis* needs to be defined. The level of analysis determines

*where*, i.e. to which extent, the analysis is conducted. This can either be the entire platform, a specific (or multiple) community or a specific member. We would like to note that – even if possible – it is not the aim of Social Collaboration Analytics to track the behavior of *a single user*. The users are only included in the scheme because they are a possible level of analysis. A further important decision is the *content type* to be included in the analysis. One or more content types can be specified. The content type refers to the applications (modules) of integrated ECS and incorporates the idea of social business documents as compound documents [22]. Additionally, the *content components* need to be specified (in accordance with the specified content type). Although, this step can be considered as optional the specification of content elements is recommended for more precise analyses. Depending on the content type, it is possible to specify one or more different content elements. The relevant data for the level of analysis is stored in the content databases (cf. Figure 2).

Next, the *action types* need to be selected. This step is particularly important and needs careful consideration. In the case of IBM Connections, there are more than 50 different action types. Each event is marked with a *timestamp*, which allows us to limit the time frame of the analysis (e.g. to a particular month or year). The relevant data for the action types and timestamps is stored in the event database (cf. Figure 2).

Where?	Which type of content?	Which content component?	Which type of action?	When?	Which filters?
Level of analysis	Content type	Content component	Action type	Time	Further filters
<ul style="list-style-type: none"> <li>Platform</li> <li>Community</li> <li>Members</li> </ul>	<ul style="list-style-type: none"> <li>Activities</li> <li>Blogs</li> <li>Bookmarks</li> <li>Files</li> <li>Forum</li> <li>Ideation Blog</li> <li>Media Gallery</li> <li>Profiles</li> <li>Wiki</li> </ul>	<ul style="list-style-type: none"> <li>Entry</li> <li>Post</li> <li>Comment</li> <li>Like</li> <li>Tag</li> <li>Status messages</li> </ul>	<ul style="list-style-type: none"> <li>Add</li> <li>Copy</li> <li>Create</li> <li>Delete</li> <li>Download</li> <li>Follow</li> <li>Read</li> <li>Recommend</li> <li>Update</li> <li>...</li> </ul>	<ul style="list-style-type: none"> <li>Timestamp</li> </ul>	<ul style="list-style-type: none"> <li>Business Unit</li> <li>Region</li> <li>Country</li> <li>Time constraints</li> <li>Week</li> <li>Month</li> <li>Quarter</li> <li>Year</li> <li>Custom</li> <li>...</li> </ul>

Figure 5: SCA Framework: Elements that can be used for the queries

Finally, *further filters* can be applied. The query can, for example, be limited to a specific business unit or a region depending on the available data in the social profiles of the authors. Depending on the required information, the query has to join data from the three available data sources (organizational data, transactional data, content data) as previously shown in Figure 2.

In the last phase of our research, we deployed our SCA framework to demonstrate its ability to answer

the questions that our practitioners provided in the focus group. We presented the results in a workshop with the same participants from the focus group and gathered feedback for evaluation purposes. Figure 6 shows one of the examples that was presented at the workshop.

month without any holidays so this month marks the peak of activity. December and January contained holidays, therefore the activity in the community is slightly lower than in the previous month. The semester finished in early February, so the activity in this month is comparable to the activity in October.

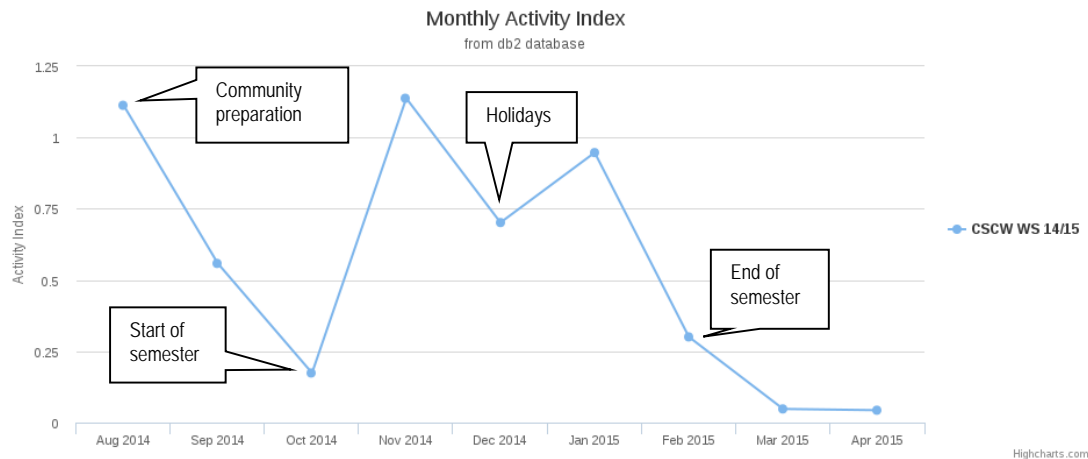


Figure 6: Activity index for a community

The aim of the exemplary query was to show the activities of all members in a given period of time (one semester) across all modules in a community that was used for teaching purposes. We decided to use the activity metric as suggested by [19] for our prototypical demonstration. The analysis was implemented in PHP and SQL. The tool Highcharts was used for the graphical representation. The elements from the framework were instantiated with the following parameters:

- Level of Analysis: Community
- Content type: All
- Content element: All
- Action type: All
- Time and further filters: August 2014 to April 2015 (period of approximately one semester)

For reasons of simplicity, we chose an example that considers all content types, content elements and action types. Figure 6 shows the graphical result, which we enriched by explanatory call-outs. The example provided is already more sophisticated than suggested by the framework because we applied an additional formula in the analysis [19].

Our example shows a community for a Bachelor class on Computer Supported Collaborative Work. In August and September, we prepared the community for the upcoming semester. These months are characterized by a high activity of internal teaching staff. The semester only started in late October, which led to a low activity in this month. November was the only

The *exemplary deployment* confirmed several aspects. Firstly, the validity of data that is recorded by the system could be verified since the generated chart matches the schedule of the lecture and represents the typical activity in a semester. Secondly, all workshop participants stated that the analysis provided in this example is useful and goes beyond the default reports that their ECS provides. In the discussion part of the workshop, the participants unanimously stated that the presentation and interpretation of the data is an important part of Social Collaboration Analytics.

## 4. Discussion

Previous research in the literature describes various methods for analyzing content and interactions in ESS. The following section aims to show the differences between Social Collaboration Analytics and previous work in this area. A theme that commonly occurs in the context of ESS is *Social Network Analytics (SNA)*. The special interest of SNA are the relationships between people in a social network and the patterns and implications that come with these relationships [28]. Methods from SNA were applied for several purposes in ESS for example for the *extraction of expertise* [29], [30], the *generation of networks* based on user interactions on documents [24], [25] or the analysis of the *effects of organizational hierarchy* on user behavior [14], [31]. Other researchers developed and/or applied methods for genre or sentiment analysis to analyze *communication* [32], [33].



Our suggested approach to Social Collaboration Analysis is complementary to previous work on Social Analytics. Our framework is meant to assist with the examination of user activity and interaction in digital workplaces (communities). The classification scheme shows the necessary data sources and the framework guides the formulation of SQL queries with the final aim of visualizing and understanding how (intensively and successfully) employees are using the Enterprise Collaboration System.

Some researchers discuss the development of metrics or key performance indicators for ESS [19], [34] for example with the aim of developing dashboards. We believe that our SCA framework and the data scheme could be used as a starting point to develop dashboard applications.

A recent publication refers to the analysis of user interactions and relationships in Enterprise Social Software as *Business Intelligence 2.0* [35]. According to Nelson it comprises enterprise integration, mobile and ubiquitous access and technical aspects such as in-memory analytics and BI as a service [36]. We agree with Nelson who points out that Business Intelligence 2.0 requires more than mere event-based collaboration analysis. It requires taking into consideration all information that a company generates, including the *transactional data* in ERP systems (the focus of traditional business intelligence), *content management systems* and the *socially-enabled collaboration systems* (ECS). It will be interesting to see how the combination of all of these business software systems will lead to an improved ad hoc real time access to information that can be used for decision taking. Companies such as IBM are announcing the era of cognitive computing [37]. We believe that our approach to Social Collaboration Analytics will increase the understanding of joint work in the digital workspace and will thus be an important source for Business Intelligence 2.0.

## 5. Conclusions and Outlook

In modern integrated ESS every type of (inter-)action is traceable. This leads to a vast amount of data that can potentially be analyzed. We classified the different type of sources of data for SCA. Our proposed SCA framework is making use of this data in a structured way. The remaining challenge for SCA is the analysis of existing sets of data by asking the right questions. The objective of our SCA framework is to support researchers and managers in finding the right questions.

Our SCA framework was derived from the databases of a specific software (IBM Connections) but we are confident that the SCA framework is technology agnostic and we will continue to work on it to make it

usable for other ESS in the market (e.g. Atlassian Confluence/Jira). Future plans include the application of process mining to ECS (an established concept in business process management [38], [39]) in order to visualize collaboration scenarios [40], [41]. Further, we will try to apply methods from text mining [42] to ECS in order to identify topics of interest.

## 6. References

- [1] S. P. Williams and P. Schubert, "Social Business Readiness Survey 2014," University of Koblenz-Landau, Working Report of the Research Group Business Software, Koblenz, 2015.
- [2] K. Riemer, S. Stieglitz, and C. Meske, "From Top to Bottom: Investigating the Changing Role of Hierarchy in Enterprise Social Networks," *Bus. Inf. Syst. Eng.*, vol. 57, no. 3, pp. 197–212, 2015.
- [3] C. Herzog, A. Richter, M. Steinhüser, U. Hoppe, and M. Koch, "Methods and metrics for measuring the success of Enterprise Social Software – what we can learn from practice and vice versa," in *21st European Conference on Information Systems*, 2013, pp. 1–12.
- [4] S. Amer-Yahia, S. Anjum, and A. Ghenai, "MAQSA: a system for social analytics on news," in *International Conference on Management of Data*, 2012, pp. 653–656.
- [5] S. P. Williams and P. Schubert, "Connecting Industry: Building and Sustaining a Practice-based Research Community," in *50th Hawaii International Conference on System Sciences (HICSS)*, 2017, pp. 1–10.
- [6] B. Kuechler and V. Vaishnavi, "Theory Development in Design Science Research: Anatomy of a Research Project," in *3rd International Conference on Design Science Research in Information Systems and Technology*, 2008, pp. 1–15.
- [7] P. Chapman, J. Clinton, R. Kerber, T. Khabaza, T. Reinartz, C. Shearer, and R. Wirth, "CRIPS-DM 1.0: Step-by-step data mining guide." CRISP-DM Consortium, p. 76, 2000.
- [8] V. Hausmann, S. P. Williams, and P. Schubert, "Developing a Framework for Web Analytics," in *International Bled Conference*, 2012, vol. 1990, pp. 1–14.
- [9] S. P. Williams, "Enterprise 2.0 and Collaborative Technologies," Working Report of the Research Group Business Software, May 2010, University of Koblenz-Landau, Koblenz, Germany, 2010.
- [10] P. Schubert and S. P. Williams, "The Concept of Social Business: Oxymoron or Sign of a Changing Work Culture?," in *26th International Bled Conference*, 2013, pp. 222–235.
- [11] K. Riemer, A. Richter, and M. Bohringer, "Enterprise microblogging," *Bus. Inf. Syst. Eng.*, vol. 2, no. 6, pp. 391–394, 2010.
- [12] S. P. Williams, V. Hausmann, C. A. Hardy, and P. Schubert, "Enterprise 2.0 Research: Meeting the Challenges of Practice," in *26th International Bled Conference*, 2013, vol. 2004, no. 2009, pp. 251–263.

- [13] P. Schubert and J. H. Glitsch, "Use Cases and Collaboration Scenarios: How employees use socially-enabled Enterprise Collaboration Systems (ECS)," *Int. J. Inf. Syst. Proj. Manag.*, vol. 4, no. 2, pp. 41–62, 2016.
- [14] S. Behrendt, A. Richter, and M. Trier, "Mixed methods analysis of enterprise social networks," *Comput. Networks*, vol. 75, pp. 560–577, 2014.
- [15] P. Schubert, N. Bjørn-Andersen, and T. Kilian, "Archetypes for Engaged Scholarship in IS," *Int. J. Inf. Syst. Manag.*, vol. 1, no. 3, pp. 219–239, 2015.
- [16] A. S. Lee, "Rigor and Relevance in MIS Research: Beyond the Approach of Positivism Alone," *MIS Q.*, vol. 23, no. 1, pp. 29–34, 1999.
- [17] N. B. Ellison and D. M. Boyd, "Sociality Through Social Network Sites," in *The Oxford Handbook of Internet Studies*, W. H. Dutton, Ed. Oxford Handbooks Online, 2013.
- [18] A. Hermida, F. Fletcher, D. Korrell, and D. Logan, "Your Friend as Editor: The Shift to the Personalized Social News Stream," in *The Future of Journalism Conference*, 2011, pp. 1–16.
- [19] N. Jeners and W. Prinz, "Metrics for Cooperative Systems," in *GROUP'14 18th International Conference on Supporting Group Work*, 2014, pp. 91–99.
- [20] G. Harden, "Knowledge sharing in the workplace: A social networking site assessment," in *45th Hawaii International Conference on System Sciences*, 2012.
- [21] K. Berger, J. Klier, M. Klier, and A. Richter, "Who is key...? - Value adding users in enterprise social networks," in *22nd European Conference on Information Systems*, 2014, pp. 1–16.
- [22] V. Hausmann and S. P. Williams, "Issues for the long-term management of Social Business Documents," *Int. J. Inf. Syst. Proj. Manag.*, no. 3, 2016.
- [23] S. Chai and M. Kim, "What makes bloggers share knowledge? An investigation on the role of trust," *Int. J. Inf. Manage.*, vol. 30, no. 5, pp. 408–415, 2010.
- [24] D. R. Millen, J. Feinberg, and B. Kerr, "Dogear: Social bookmarking in the enterprise," in *SIGCHI Conference on Human Factors in Computing Systems*, 2006, pp. 111–120.
- [25] N. S. Shami, M. Muller, and D. Millen, "Browse and discover: social file sharing in the enterprise," in *ACM Conference on Computer Supported Cooperative Work*, 2011, pp. 295–304.
- [26] F. Piller, P. Schubert, M. Koch, and K. M. Möslin, "Overcoming Mass Confusion: Collaborative Customer Co-Design in Online Communities," *J. Comput. Commun.*, vol. 10, no. 4, p. article 8, Jul. 2005.
- [27] H. M. Hasan and C. C. Pfaff, "The wiki: An environment to revolutionise employees' interaction with corporate knowledge," in *18th Australia Conference on Computer-Human Interaction: Design: Activities, Artefacts and Environments*, 2006, pp. 377–380.
- [28] S. Wasserman and K. Faust, *Social Network Analysis: Methods and Applications*. Cambridge: Cambridge University Press, 1994.
- [29] P. Nasirifard and V. Peristeras, "Expertise extracting within online shared workspaces," in *Proceedings of the WebSci'09: Society On-Line*, 2009.
- [30] P. Nasirifard, V. Peristeras, C. Hayes, and S. Decker, "Extracting and utilizing social networks from log files of shared workspaces," in *10th IFIP Working Conference on Virtual Enterprises*, 2009, vol. 307, pp. 643–650.
- [31] S. Behrendt, M. Klier, J. Klier, A. Richter, and K. Wiesneth, "The Impact of Formal Hierarchies on Enterprise Social Networking Behavior," in *36th International Conference on Information Systems*, 2015, pp. 1–19.
- [32] K. Riemer and A. Richter, "Tweet inside: Microblogging in a corporate context," in *23rd International Bled Conference*, 2010, pp. 1–17.
- [33] K. Riemer and P. Scifleet, "Enterprise Social Networking in Knowledge-intensive Work Practices: A Case Study in a Professional Service Firm," *23rd Australas. Conf. Inf. Syst.*, pp. 1–12, 2012.
- [34] B. Oti Jacques, M. Noirhomme, X. Gobert, and F. Feltz, "Cooperation Indexes to Support Workspace Awareness," in *Groupware: Design, Implementation, and Use*, no. 4154, Y. A. Dimitriadis, I. Zigurs, and E. Gómez-Sánchez, Eds. Springer Berlin Heidelberg, 2006, pp. 94–101.
- [35] S. Behrendt, A. Richter, S. Schäfer, and M. Trier, "Business Intelligence 2.0," in *Business Intelligence for New-Generation Managers*, Springer International Publishing, 2015, pp. 97–111.
- [36] G. S. Nelson, "Business Intelligence 2.0: Are we there yet?," in *SAS Global Forum 2010 Business Intelligence / Analytics Business*, 2010, pp. 1–10.
- [37] A. Sheth, "Internet of Things to Smart IoT Through Semantic, Cognitive, and Perceptual Computing," *IEEE Intell. Syst.*, vol. 31, no. 2, pp. 108–112, 2016.
- [38] W. M. P. van der Aalst, "Business Process Management: A Comprehensive Survey," *ISRN Softw. Eng.*, pp. 1–37, 2013.
- [39] W. M. P. van der Aalst, "A Decade of Business Process Management Conferences: Personal Reflections on a Developing Discipline," in *Business Process Management*, vol. 7481, A. Barros, A. Gal, and E. Kindler, Eds. Springer, 2012, pp. 1–16.
- [40] M. A. Chaves and E. R. Cordoba, "Deciphering event logs in SharePoint Server: A methodology based on process mining," in *XL Latin American Computing Conference (CLEI)*, 2014, pp. 1–12.
- [41] F. Naderipour, "Mining of ad-hoc business processes using Microsoft Sharepoint, NITRO & PROM 6.0," in *13th International Conference on Enterprise Information Systems*, 2011, pp. 413–418.
- [42] S. Debortoli, O. Müller, I. A. Junglas, and J. vom Brocke, "Text Mining for Information Systems Researchers: An Annotated Tutorial," *Commun. Assoc. Inf. Syst.*, 2016.
- [43] IBM, "Experience IBM Connections," 2016. [Online]. Available: <http://infolib.lotus.com/resources/experience/connection/s/>. [Accessed: 16-Sep-2016].